

```
[student559_10@bigdataanalytics-worker-3 ~]$ /opt/spark-2.4.8/bin/pyspark --packages org.apache.spark:spark-sql-kafka-0-10_2.11:2.4.5
Python 2.7.5 (default, Nov 16 2020, 22:23:17)
[GCC 4.8.5 20150623 (Red Hat 4.8.5-44)] on linux2
Type "help", "copyright", "credits" or "license" for more information.
Ivy Default Cache set to: /home/student559_10/.ivy2/cache
The jars for the packages stored in: /home/student559_10/.ivy2/jars
:: loading settings :: url = jar:file:/opt/spark-2.4.8/jars/ivy-2.4.0.jar!/org/apache/ivy/core/settings/ivysettings.xml
org.apache.spark#spark-sql-kafka-0-10_2.11 added as a dependency
:: resolving dependencies :: org.apache.spark#spark-submit-parent-69642e6c-602e-4b8b-b796-03f82d9307bd;1.0
  confs: [default]
    found org.apache.spark#spark-sql-kafka-0-10_2.11;2.4.5 in central
    found org.apache.kafka#kafka-clients;2.0.0 in central
    found org.lz4#lz4-java;1.4.0 in central
    found org.xerial.snappy#snappy-java;1.1.7.3 in central
    found org.slf4j#slf4j-api;1.7.16 in central
    found org.spark-project.spark#unused;1.0.0 in central
downloading https://repo1.maven.org/maven2/org/apache/spark/spark-sql-kafka-0-10_2.11/2.4.5/spark-sql-kafka-0-10_2.11-2.4.5.jar ...
[SUCCESSFUL ] org.apache.spark#spark-sql-kafka-0-10_2.11;2.4.5!spark-sql-kafka-0-10_2.11.jar (118ms)
downloading https://repo1.maven.org/maven2/org/apache/kafka/kafka-clients/2.0.0/kafka-clients-2.0.0.jar ...
[SUCCESSFUL ] org.apache.kafka#kafka-clients;2.0.0!kafka-clients.jar (116ms)
downloading https://repo1.maven.org/maven2/org/lz4/lz4-java/1.4.0/lz4-java-1.4.0.jar ...
[SUCCESSFUL ] org.lz4#lz4-java;1.4.0!lz4-java.jar (53ms)
downloading https://repo1.maven.org/maven2/org/xerial/snappy/snappy-java/1.1.7.3/snappy-java-1.1.7.3.jar ...
[SUCCESSFUL ] org.xerial.snappy#snappy-java;1.1.7.3!snappy-java.jar(bundle) (107ms)
:: resolution report :: resolve 4564ms :: artifacts dl 400ms
  :: modules in use:
    org.apache.kafka#kafka-clients;2.0.0 from central in [default]
    org.apache.spark#spark-sql-kafka-0-10_2.11;2.4.5 from central in [default]
    org.lz4#lz4-java;1.4.0 from central in [default]
    org.slf4j#slf4j-api;1.7.16 from central in [default]
    org.spark-project.spark#unused;1.0.0 from central in [default]
    org.xerial.snappy#snappy-java;1.1.7.3 from central in [default]
-----
|               | modules | artifacts |
|   conf   | number| search|dwnlded|evicted|| number|dwnlded|
-----+-----+-----+-----+-----+
| default |     6 |     4 |     4 |     0 ||     6 |     4 |
-----
:: retrieving :: org.apache.spark#spark-submit-parent-69642e6c-602e-4b8b-b796-03f82d9307bd
  confs: [default]
  4 artifacts copied, 2 already retrieved (4706kB/19ms)
22/01/29 21:09:25 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where
Setting default log level to "WARN".
To adjust logging level use sc.setLogLevel(newLevel). For SparkR, use setLogLevel(newLevel).
22/01/29 21:09:26 WARN util.Utils: Service 'SparkUI' could not bind on port 4040. Attempting port 4041.
Welcome to

  ____
 /  __ \
/   /  \
/  /_  \
/___/___\
version 2.4.8

Using Python version 2.7.5 (default, Nov 16 2020 22:23:17)
SparkSession available as 'spark'.
>>>
```

```
[student559_10@bigdataanalytics-worker-3 bin]$ /usr/hdp/current/kafka-broker/bin/kafka-console-consumer.sh --bootstrap-server bigdataanalytics-worker-3:6667 --from-beginning --max-messages 5 --topic test_lesson_3_sapr

[
  {"sepalLength": 5.1, "sepalWidth": 3.5, "petalLength": 1.4, "petalWidth": 0.2, "species": "setosa"},
  {"sepalLength": 4.9, "sepalWidth": 3.0, "petalLength": 1.4, "petalWidth": 0.2, "species": "setosa"},
  {"sepalLength": 4.7, "sepalWidth": 3.2, "petalLength": 1.3, "petalWidth": 0.2, "species": "setosa"},
  {"sepalLength": 4.6, "sepalWidth": 3.1, "petalLength": 1.5, "petalWidth": 0.2, "species": "setosa"}
]
Processed a total of 5 messages
[student559_10@bigdataanalytics-worker-3 bin]$
[student559_10@bigdataanalytics-worker-3 bin]$
```

```
>>> raw_data = spark.readStream. \
...   format("kafka"). \
...   option("kafka.bootstrap.servers", kafka_brokers). \
...   option("subscribe", "test_lesson_3_sapr"). \
...   option("startingOffsets", "earliest"). \
...   option("maxOffsetsPerTrigger", "5"). \
...   load()
```

```

>>> schema = StructType() \
...     .add("sepalLength", FloatType()) \
...     .add("sepalWidth", FloatType()) \
...     .add("petalLength", FloatType()) \
...     .add("petalWidth", FloatType()) \
...     .add("species", StringType())
>>>
>>> parsed_iris = raw_data \
...     .select(F.from_json(F.col("value").cast("String"), schema).alias("value"), "offset") \
...     .select("value.*", "offset")
>>> raw_data.printSchema()
root
|-- key: binary (nullable = true)
|-- value: binary (nullable = true)
|-- topic: string (nullable = true)
|-- partition: integer (nullable = true)
|-- offset: long (nullable = true)
|-- timestamp: timestamp (nullable = true)
|-- timestampType: integer (nullable = true)

>>> parsed_iris.printSchema()
root
|-- sepalLength: float (nullable = true)
|-- sepalWidth: float (nullable = true)
|-- petalLength: float (nullable = true)
|-- petalWidth: float (nullable = true)
|-- species: string (nullable = true)
|-- offset: long (nullable = true)

>>> █

```

```
...  
>>> out = console_output(parsed_iris, 15)  
22/01/29 21:34:00 WARN shortcircuit.DomainSocketFactory: The short-circuit  
e libhadoop cannot be loaded.
```

```
-----  
Batch: 0  
-----
```

sepalLength	sepalWidth	petalLength	petalWidth	species	offset
null	null	null	null	null	0
5.1	3.5	1.4	0.2	setosa	1
4.9	3.0	1.4	0.2	setosa	2
4.7	3.2	1.3	0.2	setosa	3
4.6	3.1	1.5	0.2	setosa	4

```
-----  
Batch: 1  
-----
```

sepalLength	sepalWidth	petalLength	petalWidth	species	offset
5.0	3.6	1.4	0.2	setosa	5
5.4	3.9	1.7	0.4	setosa	6
4.6	3.4	1.4	0.3	setosa	7
5.0	3.4	1.5	0.2	setosa	8
4.4	2.9	1.4	0.2	setosa	9

```
out.stop()  
>>> out.stop()  
>>> █
```

```

>>> def memory_sink(df, freq):
...     return df.writeStream.format("memory") \
...         .queryName("my_memory_sink_table") \
...         .trigger(processingTime='%s seconds' % freq) \
...         .start()
...
>>> stream = memory_sink(parsed_iris, 5)
>>> spark.sql("select * from my_memory_sink_table").show()
+-----+-----+-----+-----+-----+-----+
|sepalLength|sepalWidth|petalLength|petalWidth|species|offset|
+-----+-----+-----+-----+-----+-----+
|      null|      null|      null|      null|      null|      0|
|       5.1|       3.5|       1.4|       0.2|    setosa|      1|
|       4.9|       3.0|       1.4|       0.2|    setosa|      2|
|       4.7|       3.2|       1.3|       0.2|    setosa|      3|
|       4.6|       3.1|       1.5|       0.2|    setosa|      4|
|       5.0|       3.6|       1.4|       0.2|    setosa|      5|
|       5.4|       3.9|       1.7|       0.4|    setosa|      6|
|       4.6|       3.4|       1.4|       0.3|    setosa|      7|
|       5.0|       3.4|       1.5|       0.2|    setosa|      8|
|       4.4|       2.9|       1.4|       0.2|    setosa|      9|
|       4.9|       3.1|       1.5|       0.1|    setosa|     10|
|       5.4|       3.7|       1.5|       0.2|    setosa|     11|
|       4.8|       3.4|       1.6|       0.2|    setosa|     12|
|       4.8|       3.0|       1.4|       0.1|    setosa|     13|
|       4.3|       3.0|       1.1|       0.1|    setosa|     14|
|       5.8|       4.0|       1.2|       0.2|    setosa|     15|
|       5.7|       4.4|       1.5|       0.4|    setosa|     16|
|       5.4|       3.9|       1.3|       0.4|    setosa|     17|
|       5.1|       3.5|       1.4|       0.3|    setosa|     18|
|       5.7|       3.8|       1.7|       0.3|    setosa|     19|
+-----+-----+-----+-----+-----+-----+
only showing top 20 rows

>>> █

```

```
Name is already active
>>> spark.sql("select count(*) from my_memory_sink_table").show()
+-----+
|count(1)|
+-----+
|      25|
+-----+

>>> spark.sql("select count(*) from my_memory_sink_table").show()
+-----+
|count(1)|
+-----+
|      30|
+-----+

>>> spark.sql("select count(*) from my_memory_sink_table").show()
+-----+
|count(1)|
+-----+
|      30|
+-----+

>>> spark.sql("select count(*) from my_memory_sink_table").show()
+-----+
|count(1)|
+-----+
|      35|
+-----+

>>> stream.stop()
>>> █
```

```

...
>>> def file_sink(df, freq):
...     return df.writeStream.format("parquet") \
...         .trigger(processingTime='%s seconds' % freq) \
...         .option("path", "my_parquet_sink") \
...         .option("checkpointLocation", "iris_file_checkpoint") \
...         .start()
...
>>> stream = file_sink(parsed_iris, 5)
>>> █

```

```

[student559_10@bigdataanalytics-worker-3 ~]$ find / -name *.parquet
Found 4 items
drwxr-xr-x  - student559_10 student559_10      0 2022-01-28 23:15 .sparkStaging
drwxr-xr-x  - student559_10 student559_10      0 2022-01-28 14:11 for_stream
drwxr-xr-x  - student559_10 student559_10      0 2022-01-29 21:57 iris_file_checkpoint
drwxr-xr-x  - student559_10 student559_10      0 2022-01-29 21:57 my_parquet_sink
[student559_10@bigdataanalytics-worker-3 ~]$ █

```

```

drwxr-xr-x - student559_10 student559_10 0 2022-01-29 21:57 my_parquet_sink
[student559_10@bigdataanalytics-worker-3 ~]$ hdfs dfs -ls my_parquet_sink
Found 32 items
drwxr-xr-x - student559_10 student559_10 0 2022-01-29 21:59 my_parquet_sink/_spark_metadata
-rw-r--r-- 2 student559_10 student559_10 1702 2022-01-29 21:57 my_parquet_sink/part-00000-008d4b1
e-2daf-44b3-be05-44a9e029b07a-c000.snappy.parquet
-rw-r--r-- 2 student559_10 student559_10 1734 2022-01-29 21:59 my_parquet_sink/part-00000-06a403f
8-1f9e-483f-b9cf-3e4e3337c8e8-c000.snappy.parquet
-rw-r--r-- 2 student559_10 student559_10 1735 2022-01-29 21:59 my_parquet_sink/part-00000-0ff931c
2-2de2-4b32-8bbe-53d63953eb27-c000.snappy.parquet
-rw-r--r-- 2 student559_10 student559_10 1708 2022-01-29 21:59 my_parquet_sink/part-00000-12bfc9f
9-25df-4c78-b252-0c2e5e000e78-c000.snappy.parquet
-rw-r--r-- 2 student559_10 student559_10 1780 2022-01-29 21:58 my_parquet_sink/part-00000-159c13a
f-e38d-4756-93b0-35c3c677e2ba-c000.snappy.parquet
-rw-r--r-- 2 student559_10 student559_10 1744 2022-01-29 21:57 my_parquet_sink/part-00000-18c562c
5-9745-4b12-a78c-49d756fa1a45-c000.snappy.parquet
-rw-r--r-- 2 student559_10 student559_10 1748 2022-01-29 21:59 my_parquet_sink/part-00000-1a255af
6-a55a-4576-a9e2-d3cbfefb0cf2-c000.snappy.parquet
-rw-r--r-- 2 student559_10 student559_10 1744 2022-01-29 21:58 my_parquet_sink/part-00000-298ce55
c-5dcd-4f7e-b774-2f1f7b42a006-c000.snappy.parquet
-rw-r--r-- 2 student559_10 student559_10 1762 2022-01-29 21:58 my_parquet_sink/part-00000-32de69f
a-9ace-4edf-88fe-260e7fa61e0a-c000.snappy.parquet
-rw-r--r-- 2 student559_10 student559_10 1726 2022-01-29 21:58 my_parquet_sink/part-00000-4169eea
7-f9ba-4da9-95d9-c621c6e6bd07-c000.snappy.parquet
-rw-r--r-- 2 student559_10 student559_10 1735 2022-01-29 21:59 my_parquet_sink/part-00000-422bee9
5-994e-4d61-b559-0900e03ff711-c000.snappy.parquet
-rw-r--r-- 2 student559_10 student559_10 1735 2022-01-29 21:59 my_parquet_sink/part-00000-453f0d2
2-68fd-44ad-a862-373160962297-c000.snappy.parquet
-rw-r--r-- 2 student559_10 student559_10 1753 2022-01-29 21:59 my_parquet_sink/part-00000-471c434
6-24b1-465a-9de5-820855a2cf38-c000.snappy.parquet
-rw-r--r-- 2 student559_10 student559_10 1762 2022-01-29 21:58 my_parquet_sink/part-00000-473e526
3-67e4-47b3-909a-2c66f9622892-c000.snappy.parquet
-rw-r--r-- 2 student559_10 student559_10 1762 2022-01-29 21:58 my_parquet_sink/part-00000-5310fa6
4-348d-4d01-8c17-0fc4838de416-c000.snappy.parquet
-rw-r--r-- 2 student559_10 student559_10 1720 2022-01-29 21:57 my_parquet_sink/part-00000-5a05ad4
c-aebe-4954-8f0e-65158260311f-c000.snappy.parquet
-rw-r--r-- 2 student559_10 student559_10 1753 2022-01-29 21:59 my_parquet_sink/part-00000-5e231f0
6-26fc-4fe8-9cc9-5109843408e5-c000.snappy.parquet
-rw-r--r-- 2 student559_10 student559_10 1735 2022-01-29 21:59 my_parquet_sink/part-00000-65b3025
1-0243-4fa7-ac26-4b9ce3f106b9-c000.snappy.parquet
-rw-r--r-- 2 student559_10 student559_10 1741 2022-01-29 21:58 my_parquet_sink/part-00000-68c1c92
f-67e1-4342-a1ec-ba9434af034d-c000.snappy.parquet

```



```

3-2018-4c8e-ae11-44481b04d078-c000.snappy.parquet
[student559_10@bigdataanalytics-worker-3 ~]$ hdfs dfs -df -h my_parquet_sink
Filesystem                                Size      Used    Available  Use%
hdfs://bigdataanalytics-head-0.mcs.local:8020  35.4 G   14.7 G     16.1 G    42%
[student559_10@bigdataanalytics-worker-3 ~]$

```

## Kafka

```

Using Python version 2.7.5 (default, Nov 16 2020 22:23:17)
SparkSession available as 'spark'.
>>> from pyspark.sql import functions as F
>>> from pyspark.sql.types import *
>>> kafka_brokers = "bigdataanalytics-worker-3:6667"
>>> def kafka_sink(df, freq):
...     return df.selectExpr("CAST(null AS STRING) as key", "CAST(struct(*) AS STRING) as value") \
...         .writeStream \
...         .format("kafka") \
...         .trigger(processingTime='%s seconds' % freq) \
...         .option("topic", "shadrin_iris_sink") \
...         .option("kafka.bootstrap.servers", kafka_brokers) \
...         .option("checkpointLocation", "shadrin_iris_kafka_checkpoint") \
...         .start()
...
>>>

```

```
[student559_10@bigdataanalytics-worker-3 ~]$ /usr/hdp/current/kafka-broker/bin/kafka-console-consum
--topic test_iris_sink --from-beginning --bootstrap-server bigdataanalytics-worker-3:6667
[5.4, 3.4, 1.7, 0.2, setosa, 21]
[5.1, 3.3, 1.7, 0.5, setosa, 24]
[5.0, 3.4, 1.6, 0.4, setosa, 27]
[4.7, 3.2, 1.6, 0.2, setosa, 30]
[5.2, 4.1, 1.5, 0.1, setosa, 33]
[5.1, 3.7, 1.5, 0.4, setosa, 22]
[4.8, 3.4, 1.9, 0.2, setosa, 25]
[5.2, 3.5, 1.5, 0.2, setosa, 28]
[4.8, 3.1, 1.6, 0.2, setosa, 31]
[5.5, 4.2, 1.4, 0.2, setosa, 34]
[5.1, 3.8, 1.5, 0.3, setosa, 20]
[4.6, 3.6, 1.0, 0.2, setosa, 23]
[5.0, 3.0, 1.6, 0.2, setosa, 26]
[5.2, 3.4, 1.4, 0.2, setosa, 29]
[5.4, 3.4, 1.5, 0.4, setosa, 32]
```

```
>>> def kafka_sink_json(df, freq):
...     return df.selectExpr("CAST(null AS STRING) as key", "CAST(to_json(struct(*)) AS STRING) as value
... ) \
...     .writeStream \
...     .format("kafka") \
...     .trigger(processingTime='%s seconds' % freq) \
...     .option("topic", "test_iris_sink") \
...     .option("kafka.bootstrap.servers", kafka_brokers) \
...     .option("checkpointLocation", "shadrin_iris_kafka_checkpoint") \
...     .start()
...
>>> stream = kafka_sink_json(parsed_iris, 5)
>>> stream.stop()
>>>
```

```
Created topic 'test_iris_sink'.
[student559_10@bigdataanalytics-worker-3 bin]$ /usr/hdp/current/kafka-broker/bin/kafka-console-consumer.s
h --bootstrap-server bigdataanalytics-worker-3:6667 --from-beginning --topic test_iris_sink
{"sepalLength":4.9,"sepalWidth":3.0,"petalLength":1.4,"petalWidth":0.2,"species":"setosa","offset":2}
{"sepalLength":5.0,"sepalWidth":3.6,"petalLength":1.4,"petalWidth":0.2,"species":"setosa","offset":5}
{"sepalLength":5.0,"sepalWidth":3.4,"petalLength":1.5,"petalWidth":0.2,"species":"setosa","offset":8}
{"sepalLength":5.4,"sepalWidth":3.7,"petalLength":1.5,"petalWidth":0.2,"species":"setosa","offset":11}
{"sepalLength":4.3,"sepalWidth":3.0,"petalLength":1.1,"petalWidth":0.1,"species":"setosa","offset":14}
{"sepalLength":5.4,"sepalWidth":3.9,"petalLength":1.3,"petalWidth":0.4,"species":"setosa","offset":17}
{"sepalLength":5.1,"sepalWidth":3.8,"petalLength":1.5,"petalWidth":0.3,"species":"setosa","offset":20}
{"sepalLength":4.6,"sepalWidth":3.6,"petalLength":1.0,"petalWidth":0.2,"species":"setosa","offset":23}
{"sepalLength":5.0,"sepalWidth":3.0,"petalLength":1.6,"petalWidth":0.2,"species":"setosa","offset":26}
{"sepalLength":5.2,"sepalWidth":3.4,"petalLength":1.4,"petalWidth":0.2,"species":"setosa","offset":29}
{"sepalLength":5.4,"sepalWidth":3.4,"petalLength":1.5,"petalWidth":0.4,"species":"setosa","offset":32}
{"sepalLength":4.9,"sepalWidth":3.1,"petalLength":1.5,"petalWidth":0.2,"species":"setosa","offset":35}
{"sepalLength":4.9,"sepalWidth":3.6,"petalLength":1.4,"petalWidth":0.1,"species":"setosa","offset":38}
{"sepalLength":5.0,"sepalWidth":3.5,"petalLength":1.3,"petalWidth":0.3,"species":"setosa","offset":41}
{"sepalLength":5.0,"sepalWidth":3.5,"petalLength":1.6,"petalWidth":0.6,"species":"setosa","offset":44}
{"sepalLength":5.1,"sepalWidth":3.8,"petalLength":1.6,"petalWidth":0.2,"species":"setosa","offset":47}
{"sepalLength":5.0,"sepalWidth":3.3,"petalLength":1.4,"petalWidth":0.2,"species":"setosa","offset":50}
{"sepalLength":6.9,"sepalWidth":3.1,"petalLength":4.9,"petalWidth":1.5,"species":"versicolor","offset":53}
}
{"sepalLength":5.1,"sepalWidth":3.5,"petalLength":1.4,"petalWidth":0.2,"species":"setosa","offset":1}
{"sepalLength":4.6,"sepalWidth":3.1,"petalLength":1.5,"petalWidth":0.2,"species":"setosa","offset":4}
{"sepalLength":4.6,"sepalWidth":3.4,"petalLength":1.4,"petalWidth":0.3,"species":"setosa","offset":7}
{"sepalLength":4.9,"sepalWidth":3.1,"petalLength":1.5,"petalWidth":0.1,"species":"setosa","offset":10}
{"sepalLength":4.8,"sepalWidth":3.0,"petalLength":1.4,"petalWidth":0.1,"species":"setosa","offset":13}
{"sepalLength":5.7,"sepalWidth":4.4,"petalLength":1.5,"petalWidth":0.4,"species":"setosa","offset":16}
{"sepalLength":5.7,"sepalWidth":3.8,"petalLength":1.7,"petalWidth":0.3,"species":"setosa","offset":19}
{"sepalLength":5.1,"sepalWidth":3.7,"petalLength":1.5,"petalWidth":0.4,"species":"setosa","offset":22}
{"sepalLength":4.8,"sepalWidth":3.4,"petalLength":1.9,"petalWidth":0.2,"species":"setosa","offset":25}
{"sepalLength":5.2,"sepalWidth":3.5,"petalLength":1.5,"petalWidth":0.2,"species":"setosa","offset":28}
{"sepalLength":4.8,"sepalWidth":3.1,"petalLength":1.6,"petalWidth":0.2,"species":"setosa","offset":31}
{"sepalLength":5.5,"sepalWidth":4.2,"petalLength":1.4,"petalWidth":0.2,"species":"setosa","offset":34}
{"sepalLength":5.5,"sepalWidth":3.5,"petalLength":1.3,"petalWidth":0.2,"species":"setosa","offset":37}
{"sepalLength":5.1,"sepalWidth":3.4,"petalLength":1.5,"petalWidth":0.2,"species":"setosa","offset":40}
{"sepalLength":4.4,"sepalWidth":3.2,"petalLength":1.3,"petalWidth":0.2,"species":"setosa","offset":43}
```

```

>>> extended_iris = parsed_iris.withColumn("my_current_time", F.current_timestamp())
>>> extended_iris.printSchema()
root
 |-- sepalLength: float (nullable = true)
 |-- sepalWidth: float (nullable = true)
 |-- petalLength: float (nullable = true)
 |-- petalWidth: float (nullable = true)
 |-- species: string (nullable = true)
 |-- offset: long (nullable = true)
 |-- my_current_time: timestamp (nullable = false)

>>> █

```

```

>>> def foreach_batch_sink(df, freq):
...     return df \
...         .writeStream \
...         .foreachBatch(foreach_batch_function) \
...         .trigger(processingTime='%s seconds' % freq) \
...         .start()
...
>>> stream = foreach_batch_sink(extended_iris, 5)
Traceback (most recent call last):
  File "<stdin>", line 1, in <module>
  File "<stdin>", line 4, in foreach_batch_sink
NameError: global name 'foreach_batch_function' is not defined
>>>
>>> def foreach_batch_function(df, epoch_id):
...     print("starting epoch " + str(epoch_id) )
...     print("average values for batch:")
...     df.groupBy("species").avg().show()
...     print("finishing epoch " + str(epoch_id))
...

```

```

>>> stream = foreach_batch_sink(extended_iris, 5)
>>> starting epoch 0
average values for batch:
+-----+-----+-----+-----+-----+
|species| avg(sepalLength)| avg(sepalWidth)| avg(petalLength)| avg(petalWidth)| avg(offset)|
+-----+-----+-----+-----+-----+
| null | null | null | null | null | 0.0 |
| setosa|4.824999928474426|3.199999988079071|1.399999976158142|0.20000000298023224| 2.5 |
+-----+-----+-----+-----+-----+

finishing epoch 0
22/01/31 10:19:23 WARN streaming.ProcessingTimeExecutor: Current batch is falling behind. The trigger interval is 5000 milliseconds, but spent 6062 milliseconds
starting epoch 1
average values for batch:
+-----+-----+-----+-----+-----+
|species| avg(sepalLength)| avg(sepalWidth)| avg(petalLength)| avg(petalWidth)| avg(offset)|
+-----+-----+-----+-----+-----+
| setosa|4.880000019073487|3.440000057220459|1.4799999952316285|0.26000000536441803| 7.0 |
+-----+-----+-----+-----+-----+

finishing epoch 1
starting epoch 2
average values for batch:
+-----+-----+-----+-----+-----+
|species| avg(sepalLength)| avg(sepalWidth)| avg(petalLength)| avg(petalWidth)| avg(offset)|
+-----+-----+-----+-----+-----+
| setosa|4.840000152587891|3.240000009536743|1.4200000047683716|0.14000000208616256| 12.0 |
+-----+-----+-----+-----+-----+

finishing epoch 2
starting epoch 3
average values for batch:
+-----+-----+-----+-----+-----+
|species| avg(sepalLength)| avg(sepalWidth)| avg(petalLength)| avg(petalWidth)| avg(offset)|
+-----+-----+-----+-----+-----+
| setosa|5.539999961853027|3.9200000286102297|1.4200000047683716|0.3200000077486038| 17.0 |
+-----+-----+-----+-----+-----+

finishing epoch 3
starting epoch 4

```